

DOI:10.11918/202201067

# 多视图交叉一致性学习的半监督水面目标检测

冯俊健<sup>1</sup>, 李彬<sup>1</sup>, 田联房<sup>1,3</sup>, 董超<sup>2,3</sup>

(1. 华南理工大学 自动化科学与工程学院, 广州 510641; 2. 国家海洋局南海调查技术中心, 广州 510300;  
3. 自然资源部海洋环境探测技术与应用重点实验室, 广州 510300)

**摘要:**为缓解基于半监督学习的水面目标检测对有限标注样本过拟合的问题,提高无标注样本中目标提取的有效性,提出了基于多视图交叉一致性学习的半监督水面目标检测算法。首先,该算法通过数据增强的方式为训练样本生成不同的视图以丰富数据集的多样性;然后,利用所提出的多视图目标判别器为无标注样本在线生成伪标签,有助于提取无标注样本的有效信息;最后,利用所提出的多视图交叉一致性学习使同一目标实例的不同视图的输出实现交叉一致性正则化,以促进检测模型学习判别性的特征从而降低过拟合的风险。在海上和内河数据集上的实验结果表明:文中所提算法能够提高特征提取的判别性,对多类别的水面目标检测精度达到91.0%,比全监督检测算法提高了18.7%,比其他半监督检测算法提高了3.8%以上;在检测速度上,该算法达到13.1帧/s,基本满足实时性要求。所提算法通过多视图交叉一致性学习提高特征的判别性和缓解检测模型的过拟合风险,有助于提高半监督水面目标检测的性能。

**关键词:**水面目标检测;半监督学习;多视图交叉一致性学习;交叉一致性正则化;多视图目标判别器

**中图分类号:** TN911.73 **文献标志码:** A **文章编号:** 0367-6234(2023)04-0107-08

## Semi-supervised surface object detection based on multi-view cross-consistency learning

FENG Junjian<sup>1</sup>, LI Bin<sup>1</sup>, TIAN Lianfang<sup>1,3</sup>, DONG Chao<sup>2,3</sup>

(1. School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China;  
2. South China Sea Marine Survey and Technology Center, Guangzhou 510300, China; 3. Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou 510300, China)

**Abstract:** To solve the problem of overfitting with limited labeled samples for surface object detection based on semi-supervised learning and enhance the effectiveness of extracting objects from unlabeled samples, a semi-supervised surface object detection algorithm based on multi-view cross-consistency learning is proposed. First, this algorithm generates different views for training samples through data augmentations to enrich the diversity of the dataset. Then a multi-view target discriminator is advanced to generate pseudo-labels online for the unlabeled samples, extracting useful information from unlabeled samples. Finally, the multi-view cross-consistency learning is implemented to achieve cross-consistency regularization between the outputs of different views of the same instance, prompting the detection model to learn discriminant features and mitigate the risk of overfitting. The experimental results at maritime and inland rivers show that the proposed algorithm improves the discrimination of feature extraction. The detection accuracy of multi-category surface objects reaches 90.1%, 18.7% higher than the full supervised detection algorithm and over 3.8% higher than other semi-supervised detection algorithms. Regarding detection speed, the algorithm reaches 13.1 frames per second, basically meeting the real-time requirements. The algorithm through multi-view cross-consistency learning tends to improve the discrimination of features and reduce the overfitting risk of the detection model, with the performance of semi-supervised surface object detection optimized.

**Keywords:** surface object detection; semi-supervised learning; multi-view cross-consistency learning; cross-consistency regularization; multi-view target discriminator

随着水上监控的广泛普及,基于计算机视觉的水面目标检测方法有助于及时发现航道和码头附近

的可疑目标,及时对该水域进行预警,这有助于提高水上管理效率并降低人力成本。传统的水面目标

收稿日期: 2022-01-16; 录用日期: 2022-05-31; 网络首发日期: 2023-03-07

网络首发地址: <https://kns.cnki.net/kcms/detail/23.1235.T.20230306.1742.009.html>

基金项目: 广东省重点研发计划(2020B1111010002); 广东省海洋经济发展专项(GDNRC[2020]018); 2021年广东省科技专项资金(“大专项+任务清单”)(210719145863737)

作者简介: 冯俊健(1994—),男,博士研究生; 田联房(1968—),男,教授,博士生导师

通信作者: 田联房, chlfian@scut.edu.cn

检测方法大多以背景建模为主<sup>[1-2]</sup>,这些方法在光线变化、目标静止和密集等场景下很容易产生错误的检测结果,而且难以对检测到的目标类别进行有效的分类。

当前,深度学习利用其强大的学习能力,极大地提高目标检测的性能,并逐渐成为目标检测中的主流范式<sup>[3]</sup>。其中, Ren 等<sup>[4]</sup>提出区域提议网络生成可学习的目标候选区域并实现两阶段目标检测的端到端训练。王言鹏等<sup>[5]</sup>提出单次多框检测器算法提高对波浪和岸边树叶等背景因素的检测鲁棒性。然而,基于深度学习的检测模型拥有大量的可训练参数,以全监督方式有效地训练检测模型需要依赖大量的标注数据<sup>[6]</sup>。在水面目标检测中,由于目标类别的多样性和场景复杂性,人工进行大量多样性的样本标注是非常耗时且成本昂贵。

为了有效利用无标注样本和有限的标注样本,本文在之前的工作中提出了 CCAR(class-coordinate adversarial regularization)来实现半监督船舶检测,检测性能比同等条件下的全监督方式有所提高<sup>[7]</sup>。Jeong 等<sup>[8]</sup>提出一种基于一致性正则化的目标检测方法来使检测模型学习翻转不变性的特征,然而这些特征对光照变化等场景仍然存在泛化能力不足的问题。Sohn 等<sup>[9]</sup>提出一种通过自训练、样本增强和一致性正则化进行目标检测的半监督学习框架。该方法首先在标注样本上对检测模型进行预训练,接着为未标注样本生成伪标签,并进行样本增强和一致性正则化来重新训练检测模型,但是该方法依赖于鲁棒性强的检测模型来生成高质量的伪标签。然而,上述半监督检测方法在检测模型陷入过拟合时,难以在无标注样本中提取有效的目标,导致检测性能进一步下降。

针对半监督目标检测存在的问题,本文提出一种基于多视图交叉一致性学习的半监督水面目标检测算法。本文算法的核心思想是通过学习互补信息提高特征的判别性从而降低模型的过拟合风险<sup>[10-11]</sup>,该方法首先提出了交叉一致性正则化,通过数据增强为每个样本生成多个视图以产生互补信息,并使同一样本的不同视图的输出与其监督信息进行交叉学习,提高所学习特征的可分性和降低过拟合风险;其次提出多视图目标判别器,通过最优传输和多视图集成的方式分别为无标注样本在线生成分类伪标签和回归伪标签作为监督信息,有效地提高对无标注样本的利用,从而提高半监督检测模型的检测性能。

## 1 基于多视图交叉一致性学习的半监督水面目标检测算法

在半监督水面目标检测任务中,训练数据由标注数据集  $D_L = \{(d_i^L, g_i) \mid i = 1, \dots, N_L\}$  和无标注数据集  $D_U = \{d_i^U \mid i = 1, \dots, N_U\}$  组成,其中,  $d_i^L$  和  $d_i^U$  分别为输入图像,且标注数据集和无标注数据集中的目标类别标签一致。因此,本文的目的是联合有限的标注数据集和大量的无标注数据集来提升检测模型的水面目标检测性能。

本文提出的基于多视图交叉一致性学习的半监督水面目标检测的模型结构如图 1 所示,检测网络采用基于残差网络 Resnet50<sup>[12]</sup>作为特征提取器的 Faster R-CNN<sup>[4]</sup>实现二级的检测过程,并将检测问题转化为分类与回归问题。对检测网络输入一个原图像  $d^*$  ( $*$  为标注 L 或无标注 U) 和其数据增强的图像  $T(d^*)$ ,利用检测网络分别进行特征提取并得到候选区域。为了实现交叉一致性正则化,通过候选区域的匹配以保证同一候选区域的不同视图一一对应,并使无标注样本的多视图与其监督信息进行交叉学习。针对标注样本,检测网络计算其输出与其标签的监督损失;针对无标注样本,其监督信息是通过多视图目标判别器根据检测网络的输出生成的分类和回归伪标签,其中,分类伪标签通过最优传输理论缓解检测模型的有偏预测<sup>[13]</sup>,回归伪标签则通过集成多视图信息提高无标注样本中目标的定位精度。

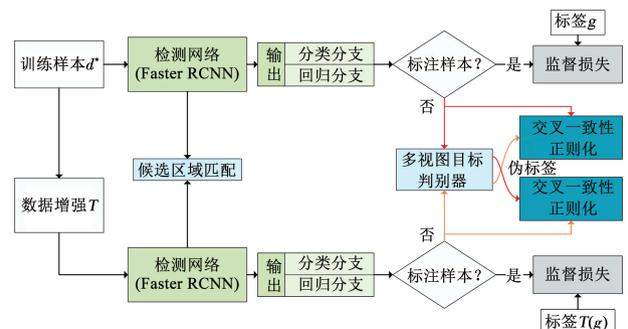


图 1 半监督水面目标检测模型结构

Fig. 1 Structure of the semi-supervised surface object detection model

### 1.1 基于交叉一致性正则化的目标函数

为了利用有限的标注样本和大量的无标注样本实现半监督训练,本文基于多视图交叉一致性学习的水面目标检测算法将检测任务分解为分类与回归任务,并提出交叉一致性正则化利用无标注样本来提高检测模型在复杂多变的水面环境下的检测性能,其目标函数为

$$L(d^L, d^U, \theta) = L_1(d^L, \theta) + \lambda R_M(d^U, T, \theta) \quad (1)$$

式中: $d^L$ 、 $d^U$ 分别为标注样本和无标注样本, $\theta$ 为检测网络的训练参数, $L_1(d^L, \theta)$ 为监督损失项, $R_M(d^U, T, \theta)$ 为交叉一致性正则项, $T$ 为数据增强变换, $\lambda$ 为权衡监督损失项和多视图交叉一致性正则项的超参数。为了简便起见,后续的公式表达将省略 $\theta$ 。

### 1.1.1 候选区域匹配

在二阶段的检测过程中,检测网络先利用RPN(region proposal network)分别将原图像 $d^*$ 数据增强图像 $T(d^*)$ 根据锚框滑动顺序提取标签为1的候选区域集合 $\mathbf{R}(d^*)$ 和 $\mathbf{R}(T(d^*))$ 并记录这些候选区域的索引号。由于索引号与特征图的锚点顺序存在固定的映射关系,对这些索引号集合进行并集计算即可得到 $d^*$ 和 $T(d^*)$ 位置匹配的候选区域集合

$\tilde{\mathbf{R}}^* = \mathbf{R}(d^*) \cup \mathbf{R}(T(d^*))$ 。其中,标注样本的候选区域标签生成根据与真实标注框的交并比标准判断;无标注样本的候选区域标签根据该候选区域的分类置信靠前 $N_R$ 的标签置1,置信度靠后 $3N_R$ 置0。合并后的候选区域的置信度为对应所有视图的平均值,并根据二分类的置信度阈值0.5划分为前景目标和背景。数据增强变换 $T$ 包括颜色抖动、高斯噪声、随机遮挡、旋转和缩放等。为了避免候选区域由于空间变换而匹配失败,结合检测网络对图像特征提取而导致的下采样,文中将数据增强变换 $T$ 对平移变换限制为 $\pm 15$ 像素,对旋转变换限制为 $\pm 3^\circ$ 。

### 1.1.2 交叉一致性正则化

交叉一致性正则化包括类别交叉一致性正则化和回归交叉一致性正则化。结合无标注样本匹配后的候选区域 $\tilde{\mathbf{R}}^U$ ,交叉一致性正则化的数学表达式为

$$R_M(d^U, T) = \frac{1}{|\tilde{\mathbf{R}}^U|} \sum_{v \in \tilde{\mathbf{R}}^U} (R_c(v, T) + R_r(v, T)) \quad (2)$$

式中: $|\tilde{\mathbf{R}}^U|$ 为候选区域数量, $v$ 为一个候选区域。 $R_c(v, T)$ 为类别交叉一致性正则化, $R_r(v, T)$ 为回归交叉一致性正则化。

对于类别交叉一致性正则化,由于同一目标的不同视图具有相同的语义信息,因此,使同一样本的不同视图所提取到的特征互相预测它们的类别输出,可以提高检测模型特征提取的泛化性,降低过拟合风险<sup>[14]</sup>。类别交叉一致性正则化的表达式为

$$R_c(v, T) = \frac{1}{N_T(N_T + 1)} \sum_{j=0}^{N_T} \sum_{k=0}^{N_T} [j \neq k] f_c(T_j(v), \hat{g}^k) \quad (3)$$

式中: $N_T$ 为数据增强变换的次数, $T_j$ 为候选区域对应的第 $j$ 种数据增强, $\hat{g}^k$ 为候选区域第 $k$ 种数据增强对应的伪标签。当 $j$ 或 $k$ 为0时,表示原图像的候选区域和对应的标签。 $[j \neq k]$ 为指示函数,当 $j \neq k$ 时为1,否则为0。 $f_c$ 为交叉熵函数。其表达式为

$$f_c(v, g) = - \sum_m p^m(g_c) \log p^m(v_c) \quad (4)$$

式中: $p^m(g_c)$ 为标签 $g$ 的第 $m$ 个类别概率, $p^m(v_c)$ 为候选区域 $v$ 的第 $m$ 个类别概率。

与类别交叉一致性正则化同理,对于回归交叉一致性正则化,同一目标根据数据增强反变换 $T^{-1}$ 将一个视图的回归输出映射到原图像的视图,并互相预测其回归输出,可以提高检测模型在回归预测中对扰动的鲁棒性。回归交叉一致性正则化的表达式为

$$R_r(v, T) = \frac{1}{N_T(N_T + 1)} \sum_{j=0}^{N_T} \sum_{k=0}^{N_T} [j \neq k] f_r(v, T_k^{-1}(\hat{g}_r^k)) \quad (5)$$

式中: $\hat{g}_r^k$ 为第 $k$ 种视图对应的回归伪标签, $T_k^{-1}$ 为第 $k$ 种数据增强的反变换, $f_c$ 为Smooth-L1函数<sup>[4]</sup>。

### 1.1.3 监督损失项

本文采用Faster RCNN<sup>[4]</sup>的二级检测网络,其监督损失项包括RPN和RCNN两部分,因此,监督损失函数的表达式为

$$L_1(d^L) = L_2(d^L) + L_3(d^L) \quad (6)$$

式中: $L_2(d^L)$ 为RPN的监督损失, $L_3(d^L)$ 为RCNN的监督损失。这两个监督损失都通过交叉熵函数和Smooth-L1函数分别计算它们的分类损失和回归损失。

## 1.2 多视图目标判别器

由于利用有限标注样本训练的模型对无标注样本进行伪标注是一种有偏估计,利用存在偏差的伪标注进行迭代训练容易加剧其模型的错误预测。本文提出多视图目标判别器,通过加入约束条件减小分类伪标签和回归伪标签的偏差。

### 1.2.1 分类伪标签

当前,分类损失一般使用交叉熵函数计算,利用检测网络将一个批次中的无标注样本生成 $N_1$ 个候选区域 $\mathbf{v}$ ,其对应的伪标签为 $\hat{\mathbf{g}}$ ,则所有无标注样本生成的候选区域的分类损失表达式为

$$E(\mathbf{v}, \hat{\mathbf{g}}) = - \frac{1}{N_1} \sum_i \sum_m p^m(\hat{\mathbf{g}}_c^i) \log p^m(\mathbf{v}_c^i) - \alpha L_c(\hat{\mathbf{g}}_c) \quad (7)$$

式中: $p^m(\hat{\mathbf{g}}_c^i)$ 为第 $i$ 个候选区域的属于第 $m$ 类别的伪标签概率, $p^m(\mathbf{v}_c^i)$ 为当前检测模型对第 $i$ 个候选区域属于第 $m$ 类别的预测概率。 $\alpha$ 为一个超参数, $L_c(\hat{\mathbf{g}}_c)$ 为熵函数,用于缓解伪标签对模型预测的偏

向,其表达式为

$$L_c(\hat{\mathbf{g}}_c) = -\frac{1}{N_1} \sum_i \sum_m p^m(\hat{\mathbf{g}}_c^i) \log p^m(\hat{\mathbf{g}}_c^i) \quad (8)$$

最小化  $E(\mathbf{v}, \hat{\mathbf{g}})$  可以求得最佳的伪标签,为了减小分类伪标签受模型偏差的影响,文中增加约束条件使一个批次内的候选区域类别分布尽可能均匀,则该问题可以转化为最优传输问题,即寻找一个最优的伪标签矩阵,将抽样概率为  $1/N_1$  的候选区域类别分布,以  $-\log p^m(\mathbf{v}_c^i)$  作为惩罚系数传输到概率为  $1/M$  的类别分布 ( $M$  为类别总数),其表达式见式(9),并通过 Sinkhorn-Knopp 算法<sup>[15]</sup> 求解。

$$\begin{cases} \hat{\mathbf{g}} = \operatorname{argmin}_{\mathbf{g} \in \mathbf{R}_+} E(\mathbf{v}, \mathbf{g}) \\ \text{s. t. } p(\mathbf{v}^i) = 1/N_1, p(\hat{\mathbf{g}}^i) = 1/M \end{cases} \quad (9)$$

### 1.2.2 回归伪标签

回归损失用于进一步优化候选区域的位置,其回归的目标值是候选区域坐标值与其有最大交并比的标注框坐标值的 4 个变换系数<sup>[4]</sup>。该回归变换表达式为:

$$\begin{aligned} t_x &= (x - x_g)/w_g, & t_y &= (y - y_g)/h_g \\ t_w &= \log(w/w_g), & t_h &= \log(h/h_g) \end{aligned} \quad (10)$$

式中:  $t_x, t_y, t_w, t_h$  分别为中心坐标、宽和高的回归目标值,  $x, y, w, h$  分别为候选区域的中心坐标、宽和高,下标  $g$  为标注框。

若利用检测网络预测的回归输出作为无标注候选区域的回归伪标签,这在后续的训练中是无法使检测网络学习到新的信息。因此,本文通过数据增强反变换集成多视图的回归伪标签,从而增加可学习的信息,降低回归伪标签的偏差。对于一个候选区域  $\mathbf{v}$  的回归伪标签与真实标签的偏差表达式为

$$e = \left| \frac{1}{N_T} \sum_{j=0}^{N_T} T_j^{-1}(\mathbf{v}_r^j) - \mathbf{g}_r \right| \quad (11)$$

式中:  $\mathbf{v}_r^j$  为候选区域第  $j$  个视图的回归伪标签,  $T_j^{-1}(\mathbf{v}_r^j)$  为目标候选区域第  $j$  个视图映射到原图像的回归输出。

设原图像的回归伪标签为  $\hat{\mathbf{g}}_r = \sum_{j=0}^{N_T} T_j^{-1}(\mathbf{v}_r^j)/N_T$ , 则其他视图的回归伪标签为  $T(\hat{\mathbf{g}}_r)$ , 其中回归伪标签包括中心坐标和其宽高,通过综合多个视图的回归伪标签可以有效减少预测偏差  $e$ , 并可以保持所有视图标注的一致性。最终利用回归伪标签近似为式(10)中的标注框,则可以得到回归的目标值。

### 1.3 基于多视图交叉一致性学习的半监督训练

本文所提的基于多视图交叉一致性学习的半监

督水面目标检测算法主要解决在半监督学习过程中,由于检测模型对有限标注样本过拟合导致无标注样本产生错误伪标签而提取无效目标的问题。该算法对标注样本和无标注样本进行数据增强并结合交叉一致性正则化,提高检测模型对遮挡、光照变化等场景的鲁棒性从而降低过拟合风险。在无标注样本的处理上,提出多视图目标判别器通过最优传输和集成的方式,在线更新它们的伪标签信息,从而提高检测模型的检测性能。该算法通过随机梯度下降法<sup>[16]</sup> 的训练方式最小化式(1)的目标函数,其训练过程如图 2 所示。

## 2 结果与分析

### 2.1 实验数据集

为了验证文中提出的基于多视图交叉一致性学习的半监督水面目标检测算法的性能,采用海上和内河的数据集 Singapore Maritime Dataset (SMD)<sup>[1]</sup> 和 SeaShips (SS)<sup>[17]</sup> 来构成大型的半监督水面目标数据集。文中选取 SMD 数据集 9 980 张图像,包括大型轮船、帆船、皮艇、浮标、渡船、小船、快艇和其他等 8 种水面目标类别,样本特点是背景简单、尺度变化、密集遮挡和光照变化等;SS 数据集由 7 000 张图片组成,包括运砂船、渔船、杂货船、集装箱船、散货船、客船等 6 种类别,样本特点是背景复杂、光照变化和目标数较少等。文中以 4:1 的比例对两个公开数据集分别划分训练集和测试集。训练集总数达到 13 584,其中包括 4 528 个标注样本合 9 056 个无标注样本,测试集总数为 3 396。

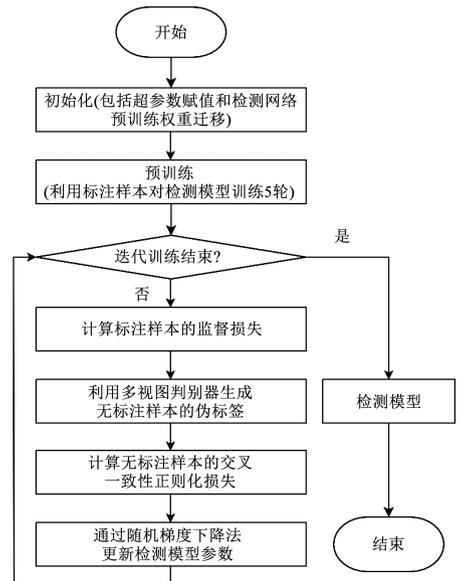


图 2 基于多视图交叉一致性学习的半监督训练过程

Fig. 2 Semi-supervised training process based on multi-view cross-consistency learning

### 2.2 实验条件及评价指标

本文方法的实验采用编程语言为 Python3.6,深度学习框架为 Pytorch1.7.1,CPU 为 Intel(R) Core(TM) i9-10900X 主频 3.70 GHz,内存为 64 GB,显卡为 NVIDIA Quadro M6000,显存为 12 GB。

本文方法的参数设置情况:通过实验发现,当式(1)的  $\lambda \leq 0.1$  能达到较好的检测性能,但  $\lambda$  太小不利于训练收敛,因此  $\lambda = 0.1$ ;同理,式(7)的  $\alpha = 0.1$ ;学习率为 0.001;训练批次设置为 4,包括一个标注样本和一个无标注样本,以及它们对应的数据增强样本;数据增强变换  $T$  在以下 5 种中随机切换:颜色抖动、高斯噪声、随机遮挡、旋转和缩放;迭代轮次为 30;为了减少 GPU 内存的占用,在训练阶段,图像输入尺寸调整为  $600 \times 1067$ ,各样本的候选区域随机保留 128 个,其中前景目标数与背景目标数比例为 1:3,因此  $N_R = 32$ ,在检测模型推理阶段,则选择候选区域置信度最高的 128 个。

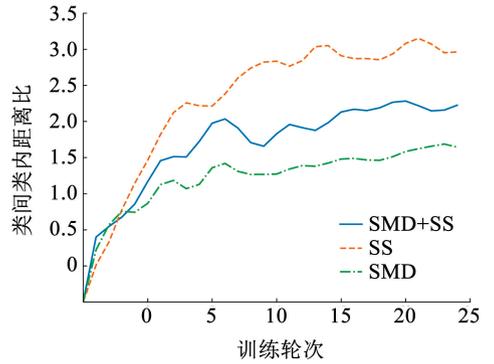
本文的评价指标有:1)类间类内距离的比值,评估特征聚类性能;2)mAP(mean average precision),评估水面目标检测精度;3)帧率(Frame per second,FPS),评估水面目标检测实时性。

### 2.3 所提算法的有效性验证实验

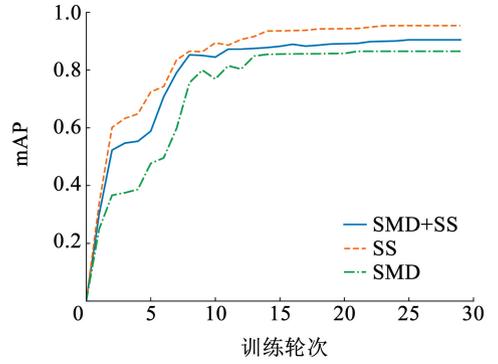
为了验证所提算法能够通过多视图交叉一致性学习能够有效提高特征的判别性和降低过拟合风险,本文通过特征类间类内距离的比值以及特征的降维可视化对水面目标检测性能进行评估。该实验将测试集中的样本输入已训练好的检测模型中,输出各候选区域的特征向量并将其赋予与标注框交叉比大于 0.5 的类别标签。

图 3 为特征可分性与检测性能的关系图,其中不同的曲线代表不同的数据集。可以看出,随着训练轮次的增加,图 3(a)中的特征类间类内距离比呈现增长的趋势,与之对应的图 3(b)中的 mAP 值也在增长。当训练轮次 < 10 时,检测模型快速地学习样本的简单信息,3 条曲线的特征类间类内距离比和 mAP 同时快速地增长。在  $10 < \text{训练轮次} < 20$  时,检测模型不断地调整特征的提取方式以更好地提高泛化性,导致特征类间类内距离比开始波动式地增长。

在训练轮次 > 20 后,检测模型在困难样本中尝试调整参数,导致特征类间类内距离比值小范围波动,此时检测模型的 mAP 趋于收敛。结合图 3(a)、3(b)可以进一步分析,极大化特征类间距离而最小化类内距离有助于提高特征的判别性并降低过拟合风险,从而提高检测模型的检测性能。



(a) 不同训练轮次的特征类间类内距离比



(b) 不同训练轮次的 mAP

图 3 特征判别性与检测性能的关系

Fig. 3 Relationship between feature separability and detection performance

图 4 为通过 t-SNE 降维的特征可视化图,其中图 4(a) 为全监督算法<sup>[4]</sup>所生成的特征可视图,图 4(b) 为本文所提算法所生成的特征可视图。可以看出,图 4(a) 由于一些类别簇与其他类别簇存在大面积的混合,在特征空间的决策边界很容易产生大量误判。当这种情况发生在半监督学习的时候,无标注样本会由于错误的预测,而扩大错误的决策边界,导致检测模型的检测性能下降。图 4(b) 通过多视图交叉一致性学习,使检测模型学习到判别性,减少特征空间中不同类别簇的混合,有助于提高半监督目标检测的性能。

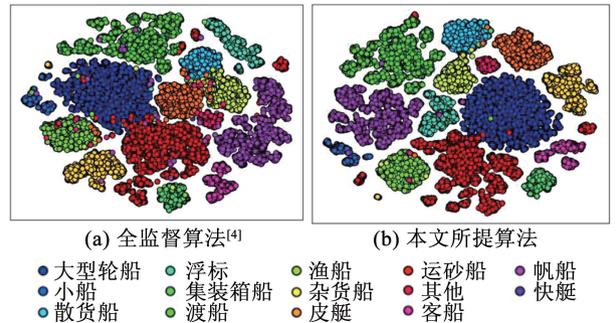


图 4 降维特征可视化

Fig. 4 Feature visualization for dimensionality reduction

### 2.4 所提算法与其他算法的对比实验

本文提出基于多视图交叉一致性学习的半监督水面目标检测算法,利用交叉一致性正则化使检测网络通过互补信息学习判别性的特征,并通过多视图目标判别器在线生成伪标签,有效地利用无标注样本进行半监督水面目标检测。为了进一步验证该算法的有效,文中与其他 4 种方法进行对比实验,包括全监督学习方法<sup>[4]</sup>和半监督学习方法<sup>[7-9]</sup>。同时,文中将交叉一致性正则化同时应用在标注样本和无标注样本上,其实验通过上标“L,U”进行标识。

表 1、2 分别为不同算法在 SS 测试集和 SMD 测试集下的 mAP。可以看出,文中所提算法在内河和海上的测试集中均获得最高的 mAP 值,分别为 95.5% 和 87.7%。而将标注样本也输入到交叉一致性正则化中可以进一步降低标注样本的过拟合风险,提高 SS 和 SMD 测试集的 mAP 值,分别为 95.7% 和 88.0%。全监督学习算法仅能利用标注样本的信息,其 mAP 值最低,分别为 80.7% 和 66.1%。半监督学习方法利用了无标注样本的信息,均不同程度地对 mAP 有所提高。其中,CCAR<sup>[7]</sup>通过类别坐标一致性正则化使半监督水面目标检测的 mAP 值分别为 84.2% 和 77.4%;Jeong 等<sup>[8]</sup>使用 Single-Shot-Detector 的网络结构一定程度上提高了小船等小目标的检测精度,它在两个测试集中的 mAP 值分别为 86.7% 和 80.2%。Sohn 等<sup>[9]</sup>通过离

线伪标注的形式获得 mAP 值分别为 92.1% 和 83.6%。表 3 为不同方法在两个测试集中的综合检测性能。可以看到,本文算法帧率为 13.1,基本满足实时性的要求。本文算法的 mAP 整体上比全监督学习算法高 18.7%,说明无标注样本的有效利用可以极大提高检测的精度。本文算法的 mAP 整体上比其他半监督算法<sup>[7-9]</sup>分别高 10.7%、8.0% 和 3.8%,说明多视图交叉一致性学习能够结合多视图的信息提高检测模型的特征判别性,并利用目标判别器在线不断更新伪标签信息,提高半监督水面目标检测的性能。值得注意的是,标注样本和无标注样本共同进行交叉一致性正则化可以进一步降低模型的过拟合风险并提高对无标注样本的有效利用,然而标注样本的优化主要受监督损失的影响,因此对 mAP 贡献的幅度不会太大。

表 1 不同方法在 SS 数据集下的平均检测精度

Tab.1 The mAP of different algorithms on SS dataset

方法	mAP/%	客船	运砂船	杂货船	渔船	集装箱船	散货船
文献[4]	80.7	73.1	79.8	88.7	70.0	88.4	83.9
文献[7]	84.2	80.7	82.2	84.6	81.6	90.3	85.8
文献[8]	86.7	86.3	82.7	88.7	83.8	91.2	87.6
文献[9]	92.1	93.2	87.1	91.2	90.9	95.4	94.7
本文	<b>95.5</b>	<b>96.1</b>	<b>96.8</b>	<b>94.3</b>	<b>93.7</b>	<b>96.7</b>	<b>95.4</b>
本文 <sup>L,U</sup>	<b>95.7</b>	<b>96.3</b>	<b>96.8</b>	<b>95.0</b>	<b>93.9</b>	<b>96.7</b>	<b>95.5</b>

注:L,U 为交叉一致性正则化同时应用在标注样本和无标注样本上。

表 2 不同方法在 SMD 数据集下的平均检测精度

Tab.2 The mAP of different algorithms on SMD dataset

方法	mAP/%	帆船	皮艇	大型轮船	浮标	渡船	其他	小船	快艇
文献[4]	66.1	87.9	58.9	87.2	62.2	67.6	61.3	60.8	42.5
文献[7]	77.4	84.5	69.4	88.9	90.4	79.7	75.3	65.3	65.9
文献[8]	80.2	90.2	81.1	85.7	89.9	81.3	63.2	71.2	78.9
文献[9]	83.6	97.5	80.9	92.7	92.8	85.2	76.4	68.7	74.7
本文	<b>87.7</b>	<b>98.2</b>	<b>82.4</b>	<b>96.2</b>	<b>95.2</b>	<b>96.0</b>	<b>80.2</b>	<b>69.5</b>	<b>83.7</b>
本文 <sup>L,U</sup>	<b>88.0</b>	<b>98.2</b>	<b>82.8</b>	<b>96.5</b>	<b>95.4</b>	<b>96.0</b>	<b>80.3</b>	<b>70.4</b>	<b>84.0</b>

注:L,U 为交叉一致性正则化同时应用在标注样本和无标注样本上。

表 3 不同算法的检测性能

Tab.3 Detection performance of different algorithms

方法	mAP/%	帧率
文献[4]	72.3	13.2
文献[7]	80.3	5.4
文献[8]	83.0	16.8
文献[9]	87.2	12.8
本文	<b>91.0</b>	<b>13.1</b>
本文 <sup>L,U</sup>	<b>91.3</b>	-

注:L,U 为交叉一致性正则化同时应用在标注样本和无标注样本上。

图 5 为所提算法在光照变化、模糊、尺度变化和遮挡等复杂场景中的船舶检测结果图。可以看到,在光照变化和模糊场景中,所提算法依然具备良好的检测性能。在尺度变化场景中,图 5(3-a)图片来自 SS 数据集,所提算法通过多视图交叉一致性学习能够该样本集中未标注的“快艇”类别。同样地,在遮挡场景中,图 5(4-a)中展示了文中模型能够检测到该样本所在的 SS 数据集中未标注的“浮标”类别。这进一步说明了本文所提算法能够有效地降低检测模型的过拟合风险,并提高对水面目标的判别性能。



图5 本文算法的可视化检测结果

Fig.5 Visualization of detection results of the proposed algorithm

### 3 结论

1)为了降低检测模型的过拟合风险和提高无标注样本的目标提取准确性,本文提出一种多视图交叉一致性学习的半监督水面目标检测算法,通过所提的交叉一致性正则化和多视图目标判别器来分别提高所学习特征的判别性和在线伪标签的准确性。

2)在 SS 和 SMD 数据集上分别取得 95.5% 和 87.7% 的 mAP 值,优于基准的全监督检测算法和其他半监督检测算法,并且检测帧率达到 13.1,基本满足实时性要求。

3)在与训练样本不相似的场景中,如利用地面采集的训练样本对空中采集的样本进行测试,本文算法的检测性能将下降。在未来工作中,将研究如

何结合迁移学习中的域自适应方法,提高检测模型的泛化能力,进一步提高对无标注样本的有效利用。

### 参考文献

[1] PRASAD D K, RAJAN D, RACHMAWATI L, et al. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(8): 1993. DOI: 10.1109/TITS.2016.2634580

[2] PRASAD D K, PRASATH C K, RAJAN D, et al. Object Detection in a maritime environment: Performance evaluation of background subtraction methods [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(5): 1787. DOI: 10.1109/TITS.2018.2836399

[3] WU Xiongwei, SAHOO D, HOI S C H. Recent advances in deep learning for object detection[J]. Neurocomputing, 2020, 396: 39. DOI: 10.1016/j.neucom.2020.01.085

- [4] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137. DOI: 10.1109/TPAMI.2016.2577031.
- [5] 王言鹏, 杨飏, 姚远. 用于内河船舶目标检测的单次多框检测器算法[J]. 哈尔滨工程大学学报, 2019, 40(7): 1258  
WANG Yanpeng, YANG Yang, YAO Yuan. Single shot multibox detector for ships detection in inland waterway[J]. Journal of Harbin Engineering University, 2019, 40(7): 1258. DOI: 10.11990/jheu.201805057
- [6] OLIVER A, ODENA A, RAFFEL C, et al. Realistic evaluation of deep semi-supervised learning algorithms[C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: MIT Press, 2018: 3239. DOI: 10.5555/3327144.3327244
- [7] FENG Junjian, LI Bin, TIAN Lianfang. Semi-supervised learning for ship detection based on class-coordinate adversarial regularization [M]// Developments of Artificial Intelligence Technologies in Computation and Robotics. Cologne: WORLD SCIENTIFIC, 2020: 882. DOI: 10.1142/9789811223334\_0106
- [8] JEONG J, LEE S, KIM J, et al. Consistency-based semi-supervised learning for object detection [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: MIT Press, 2019: 10758. DOI: 10.5555/3454287.3455252
- [9] SOHN K, ZHANG Zizhao, LI Chunliang, et al. A simple semi-supervised learning framework for object detection [J]. arXiv, 2020: 1. DOI: 10.48550/arXiv.2005.04757
- [10] FINI E, SANGINETO E, LATHUILIÈRE S, et al. A unified objective for novel class discovery[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC: IEEE, 2021: 9284. DOI: 10.1109/ICCV48922.2021.00915
- [11] LI Yingming, YANG Ming, ZHANG Zhongfei. A survey of multi-view representation learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(10): 1863. DOI: 10.1109/TKDE.2018.2872063
- [12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas Nevada: IEEE, 2016: 770. DOI: 10.1109/CVPR.2016.90
- [13] ASANO Y M, RUPPRECHT C, VEDALDI A. Self-labelling via simultaneous clustering and representation learning [C]// Proceedings of the International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020: 1
- [14] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments[C]// Proceedings of the 34th Conference on Neural Information Processing Systems. Virtual: Curran Associates Inc., 2020: 9912
- [15] CUTURI M. Sinkhorn distances: Lightspeed Computation of Optimal Transport [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. Nevada: Curran Associates Inc., 2013: 2292. DOI: 10.5555/2999792.2999868
- [16] LOSHCHELOV I, HUTTER F. SGDR: Stochastic gradient descent with warm restarts[C]// Published as a Conference Paper at ICLR 2017. Toulon: OpenReview.net, 2017: 1
- [17] SHAO Zhenfeng, WU Wenjing, WANG Zhongyuan, et al. SeaShips: A large-scale precisely annotated dataset for ship detection[J]. IEEE Transactions on Multimedia, 2018, 20(10): 2593. DOI: 10.1109/TMM.2018.2865686

(编辑 张红)